

Uma análise dos casos recepcionados pelo Judiciário brasileiro

Carolina Djovana da S. Freire

Instituto Politécnico de Coimbra, Coimbra Business School, ISCAC, Coimbra, Portugal

a2021103980@alumni.iscac.pt

Resumo – O presente artigo tem como finalidade apresentar uma breve análise do número de casos recebidos e sob incumbência do Poder Judiciário Brasileiro, um tópico de grande relevo e essencial para o desenvolvimento do país, como um dos Três Poderes da República. Tentaremos demonstrar ao longo do presente trabalho, as técnicas e modelos utilizados, que poderão ser um contributo social ao país e as dificuldades enfrentadas no seu desenvolver, na medida em que se deve considerar o contexto jurídico de todos os dados analisados. Na realização deste estudo foi usado o Google Colab, através da linguagem de programação Python.

Keywords—judiciário – brasil – métodos de classificação

I. INTRODUÇÃO

De início é importante estabelecer que este trabalho busca utilizar métodos de classificação para analisar se há relação entre os números de novos casos recebidos com o número de casos arquivados definitivamente no Judiciário brasileiro - que ocorre nas situações em que não há expectativa de prosseguimento do processo, ou seja, o processo chegou ao seu fim definitivo.

Para isso, utilizamos da base de dados obtida por meio de acesso público ao banco de dados do “Sistema de Estatísticas do Poder Judiciário” [1], principal base estatística da matéria no país e que busca a ampliação da transparência e *accountability* do Judiciário brasileiro. Esse banco de dados utiliza-se da Base Nacional de Dados do Poder Judiciário – DataJud, que é responsável pelo armazenamento centralizado dos dados e metadados processuais relativos a todos os processos físicos ou eletrônicos, públicos ou sigilosos dos Tribunais.

Esse tipo de análise se faz importante pelo grande volume de ações em andamento no Brasil. Atualmente, o Brasil conta com cerca de 81,4 milhões de processos em tramitação, aguardando alguma solução definitiva, de acordo com os dados fornecidos pelo relatório anual do Conselho Nacional de Justiça (CNJ) [2]. Diante de tamanho portfólio, se faz indispensável o controle e a constante análise e acompanhamento desses processos[3].

Além disso, vale comentar sobre a organização do DataFrame escolhido: As linhas representam cada um dos ramos judiciais do Brasil, desde 2009 até 2022, i) a Estadual, que é competente para processar e julgar qualquer causa que não esteja sujeita à Justiça Federal comum, do Trabalho, Eleitoral e Militar (ex. Tribunal de Justiça do Estado de São Paulo); ii) a Federal, que tem como missão a pacificação dos conflitos que envolvem os cidadãos e a Administração Pública

Federal, em diversas áreas (ex. Tribunal Regional Federal da 4ª Região); e iii) os Tribunais Superiores, que são 5, o Federal, o de Justiça, o do Trabalho, o Eleitoral, e o Militar (ex. o Supremo Tribunal Federal é o órgão máximo de todo o Judiciário, sendo responsável pela guarda da Constituição Federal).

Dessa forma, não se poderia realizar qualquer análise considerando que todos os Tribunais atuem da mesma forma e representem os mesmos acontecimentos. A forma simplificada de se levar esse problema é a divisão da Justiça em 3 fases: i) a 1ª instância, que é a porta de entrada do Poder Judiciário, por onde se inicia a maior parte dos processos que são analisados e julgados por um único juiz; ii) a 2ª instância, que é aquela à qual se recorre quando se pretende modificar decisão ou sentença do juízo de primeira instância, na qual o processo é julgado por uma turma de 'juízes', os desembargadores e iii) os Tribunais Superiores, que, como acima destacado, possuem competências especializadas e são compostos por Ministros.

II. METODOLOGIA

A. Importação das Bibliotecas e dos Dados

Damos início ao projeto realizando a importação das bibliotecas e dos dados que serão utilizados na linguagem de programação Python, em Google Colab. A base de dados importada se encontra em formato csv e contém os dados sobre a Justiça Nacional do Brasil desde 2009 até 2022, sendo que não estão incluídos os dados pormenorizados da Justiça Federal, que é o setor do Poder Judiciário que possui competência para julgar ações nas quais a União, suas autarquias, fundações e empresas públicas figurem como autoras ou rés, e ainda outros.

Considerando que o projeto foi desenvolvido no Google Colab, a melhor forma para se importar a base foi através do *google drive*, o que permite o compartilhamento de todos os arquivos pertinentes com as partes interessadas.

Passamos então a inspeção inicial da base de dados e como era esperado, não se encontraram colunas duplicadas. Os formatos e valores encontrados foram os seguintes:

```
df.info()
#Informação dos formatos e valores

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1189 entries, 0 to 1188
Columns: 2289 entries, chave to kmvara
dtypes: int64(5), object(2284)
memory usage: 20.8+ MB
```

Fig 1. Extraído do notebook

Acontece que, com tantas colunas e linhas, a análise fica dificultada para o fim que se propõe, a solução é a exclusão de diversas colunas e linhas desinteressantes para o presente projeto. A escolha dos dados a serem excluídas foi feita manualmente, mediante análise do arquivo em seu formato csv, no qual foi possível verificar o significado de cada nomenclatura dado às colunas e linhas. Assim, após tal correção, os formatos e valores que permaneceram passaram a nos permitir um melhor estudo:

```
[ ] df3.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 386 entries, 0 to 1163
Columns: 114 entries, justica to tbaixjudje
dtypes: int64(2), object(112)
memory usage: 346.8+ KB
```

Fig 2. Extraído do notebook

B. Tratamento dos dados

Neste DataFrame diversos dados constam como 'nd', mas na realidade estão em falta. Isso faz com que o DataFrame indique falsamente que não constem valores em falta em algumas colunas. Passamos então para a correção desse problema, substituindo os valores s 'nd' por 'NaN', assim temos enfim os reais valores em falta na DataFrame:

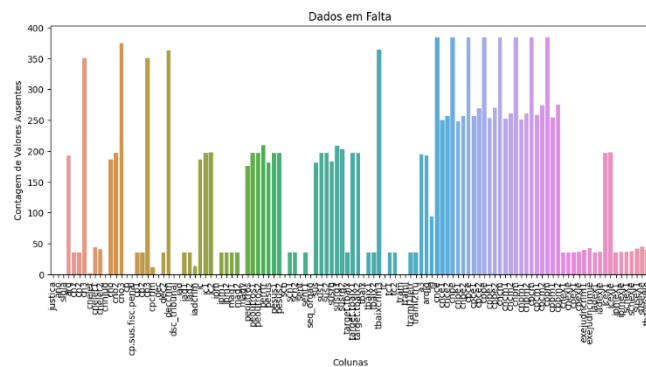


Fig 3. Extraído do notebook

Conforme se verifica do gráfico acima, existe uma quantidade significativa de dados em falta em diversas colunas do DataFrame, o que pode vir a interferir nas análises que se pretende realizar e a forma como o gráfico está representado já nos mostra que pode haver alguma relação entre os dados omissos.

Uma hipótese é que os dados em falta estejam relacionados com o ano de coleta dos dados. Neste DataFrame, como se esperaria, o número de dados em falta tende a diminuir com a passagem dos anos, salvo algumas exceções. Assim, não aparenta haver conexão entre a ausência de dados e o ano de armazenamento.

A partir disso precisamos avaliar as variáveis do DataFrame de acordo com os seus respectivos significados e limitações inerentes. Assim, passamos a análise das 30 siglas que possuem mais dados em falta e com a disponibilização

dos significados de cada coluna é possível perceber que os dados em falta fazem sentido do ponto de vista jurídico. Os dados em falta envolvem os Tribunais Superiores, casos específicos envolvendo determinadas partes e limitantes, como 1º e 2º grau de justiça. Não se poderia, portanto, realizar a previsão de todos os dados em falta sem discriminação, já que alguns dos dados em falta são decorrências lógicas do sistema de justiça que se analisa. A melhor solução é, portanto, o preenchimento dos dados em falta pelo valor '0'.

C. Visualização e Identificação de Outliers

Aqui, começamos a analisar o número total de casos novos recebidos no 1º e 2º Grau de jurisdição e nos Tribunais Superiores e o número geral de processos que são arquivados definitivamente, que no DataFrame estão definidos como 'cn1', 'cn2', 'cn3' e 'arq'.

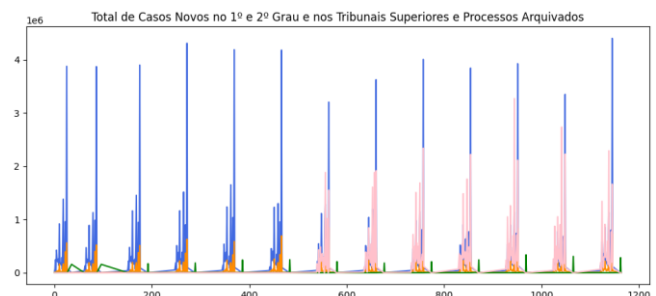


Fig 4. Extraído do notebook

De acordo com o gráfico acima, pode-se concluir que o 1º Grau de Jurisdição é o que mais recebe casos novos, sendo possível presumir também que o número de casos arquivados não era considerado nos levantamentos anteriores ao ano de 2014, pois é impossível que esses valores sejam 0, vez que é lógica jurídica que processos sejam arquivados frequentemente.

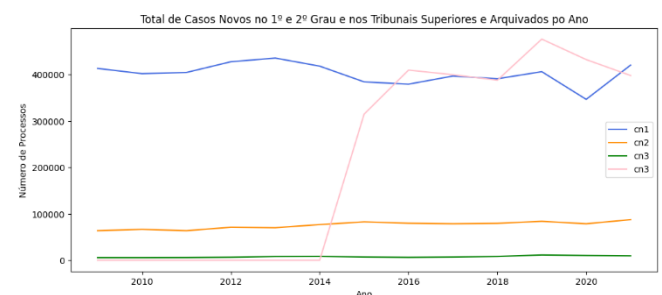


Fig 5. Extraído do notebook

Além disso, podemos visualizar que o número de casos recebidos pelos Tribunais Superiores é infinitamente inferior ao número de casos recebidos nos Tribunais Estaduais.

Passamos em seguida a verificação e identificação de outliers no DataFrame, sendo possível identificar que o estado de São Paulo é responsável pelos outliers da justiça Estadual, o que faz sentido, já que São Paulo é o estado mais populoso do país e recebe uma quantidade incomum de casos anualmente [4].

Já nos Tribunais Superiores, o que possui mais casos novos é o do Trabalho, o que é uma surpresa, já que a

suposição da maioria seria o STJ, pelo tipo de trabalho que desenvolve, sendo o Tribunal responsável por resolver conflitos de competência entre Tribunais Estaduais.

Em relação a última variável, os Tribunais que mais arquivaram foram o TJSP e o TJRJ (Tribunal de Justiça do Estado do Rio de Janeiro), o que também está em harmonia considerando que são os estados mais populosos do Brasil [5].

D. Classificação

No contexto da classificação, utilizamos as variáveis 'cn1', 'cn2' e 'cn3' para prever e classificar os valores de 'arq' e ao fim, verificar se é possível afirmar se há relação entre os números novos de casos recebidos com o número de casos arquivados definitivamente.

Para tanto, primeiramente utilizamos o método 'HistGradientBoostingClassifier', que é um classificador de machine learning da biblioteca scikit-learn. É uma implementação de aumento de gradiente usando histogramas como uma representação dos recursos de entrada durante o processo de treinamento. Este classificador é conhecido por sua eficiência e escalabilidade, principalmente para conjuntos de dados com grande número de amostras ou características.

Utilizando esse método, percebe-se que o desempenho do método ficou pouco acima dos 50%, podendo-se dizer que há certa relação entre os casos novos recebidos em cada instância e o número de casos arquivados definitivamente, mas que essa relação não chega perto de ser absoluta:

```
[ ] características = ['cn1', 'cn2', 'cn3']
X = pda[características]
y = pda['arq']

#Divisão do conjunto de dados em treino e teste
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

#Criar e treinar o modelo de classificação HistGradientBoosting
model = HistGradientBoostingClassifier()
model.fit(X_train, y_train)

#Previsões no conjunto de teste
y_pred = model.predict(X_test)

#Avaliação da acurácia do modelo
accuracy = accuracy_score(y_test, y_pred)
print("Acurácia:", accuracy)

#O desempenho do teste ficou pouco acima dos 50%, podendo-se dizer que há certa relação entre os
Acurácia: 0.5769230769230769
```

Fig 6. Extraído do notebook

Em seguida, considerando que o método utilizado acima utiliza de funciona treinando iterativamente uma sequência de árvores de decisão, onde cada árvore subsequente tenta corrigir os erros cometidos pelas árvores anteriores, penso que seja válida a análise de outro método, um mais comumente usando, o 'KNeighborsClassifier'.

Utilizando esse método, vemos que esse é mais assertivo quando utilizamos o número ideal de 4 neighbors, quando se tem desempenho de cerca de 62%, um melhor resultado do que o anterior:

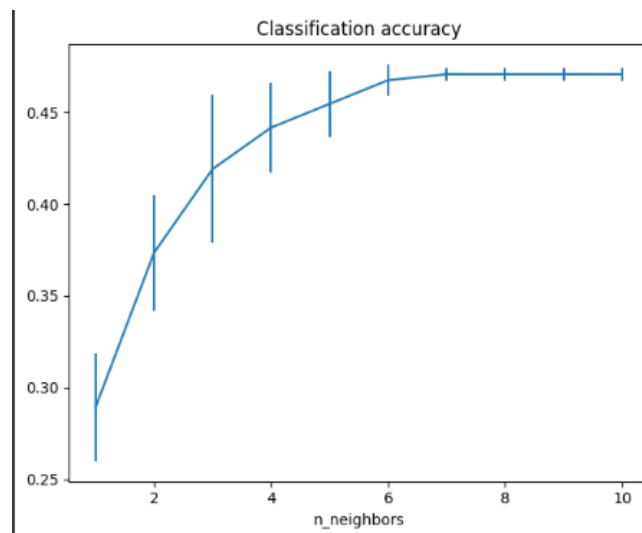


Fig 7. Extraído do notebook

De tudo, têm-se que os extensos dados do Judiciário brasileiro são extremamente ricos e passíveis de inúmeras análises, como já se é feito pelos órgãos estatísticos do país [6], o que certamente não impede e outros tipos de pesquisas, tal como essa, serem feitas a enriquecer o arsenal informacional em relação ao tema. No presente caso vemos que o grande número de dados em falta na base pública disponibilizada acaba por dificultar as análises pretendidas, assim como o grande volume de dados disponibilizados sem filtro e em um só arquivo. Apesar disso, é interessante e importante contemplar as peculiaridades do Sistema Jurídico e as tendências do setor, principalmente considerando que o Brasil tem maior número de advogados por habitantes do mundo [7], o que certamente influencia na quantidade de litígios judicializados anualmente. Em relação aos métodos utilizados, esse foi outro desafio, especialmente na sua escolha, quando se considerou qualidade e utilidade no trabalho. Contudo, tendo que no melhor dos casos temos 62% de chance de que o número de casos arquivados tenha relação com o número de casos recepcionados, é de se pensar que esse número não é robusto o suficiente para se considerar com definitiva certeza que quanto mais casos recepcionados, mais casos arquivados.

III. CONCLUSÕES

Nesta análise, a partir de um conjunto de dados reais e atuais, analisamos o número de novas ações recepcionadas pelas 3 instâncias do Poder Judiciário brasileiro. Sem dúvida um dos maiores desafios do projeto foi a análise inicial e a necessidade de realização de diversas correções da DataFrame, vez que o desafio recai sob a impossibilidade de se aplicar soluções gerais e uniformes para todos os casos, pelo decurso lógico dos significados de cada linha e coluna do DataFrame. Não obstante, foi possível perceber que não é necessariamente o grande volume de novos casos que faz com que os casos antigos sejam arquivados, pelo contrário, aparentemente o Judiciário corre seu curso sem pressa, na velocidade natural de seu trabalho, sendo que os casos são julgados e — consequentemente — arquivados, independentemente do grande número de casos que recepciona. Assim, o que podemos pensar é que o número de processos passe a se acumular no decorrer dos anos, não

havendo previsão para diminuição do trabalho desse Poder, mas o aumento.

REFERÊNCIAS

[1] <https://www.cnj.jus.br/base-de-dados>

[2] Justiça em Números 2023 / Conselho Nacional de Justiça. – Brasília: CNJ, 2023. 326 p.: il. ISBN: 978-65-5972-116-0. Disponível em: <https://www.cnj.jus.br/wp-content/uploads/2023/08/justica-em-numeros-2023.pdf>

[3] <https://www.jota.info/opiniao-e-analise/artigos/dados-juridicos-e-indicadores-de-sucesso-para-a-tomada-de-decisoes-17062020>

[4] <https://www.ibge.gov.br/cidades-e-estados/sp.html>

[5] <https://www.ibge.gov.br/cidades-e-estados/rj.html>

[6] <https://www.cnj.jus.br/pesquisas-judiciarias/justica-em-numeros/>

[7] <https://www.oab.org.br/noticia/59992/brasil-tem-1-advogado-a-cada-164-habitantes-cfoab-se-preocupa-com-qualidade-dos-cursos-juridicos>